

# DATA VISUALIZATION: TOWARD VISUALIZATION AWARENESS IN OVERLAPPED DOTS MANAGEMENT TECHNIQUES

<sup>1</sup>WAN AZZURA WAN RAMLI, <sup>2</sup>ZAINURA IDRUS, <sup>3</sup>ZANARIAH IDRUS,  
<sup>4</sup>SITI NURBAYA ISMAIL

<sup>1,2,5</sup> Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

<sup>3,4</sup> Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah, Malaysia

<sup>1</sup>azzurawpkl@gmail.com, <sup>2</sup>zainura@tmsk.uitm.edu.my, <sup>3</sup>zanaidrus@kedah.uitm.edu.my,

<sup>4</sup>sitinurbaya@kedah.uitm.edu.my

## ABSTRACT

Today, lead business, government and academic leader used data visualization to make better decision for the specific problem. Data visualization is used to visualize and analyze complex data in various fields of knowledge. However, in dot-plot based visualization, over plot or overlap dot occur when visualizing huge data. The overlapped dots, is a scenario where dots are plotted on top of each other. When this occurs, they create convoluted and confusing visual and data presented are not accurate. There are various techniques to manage overlapped dots. However, the discussion is on the individual technique and grouping them into their similarities is rarely discussed. Moreover, the discussion hardly focuses on the impact of such techniques to visualization awareness. The awareness is where users are aware that the original data have gone through some pre-processed stage thus, what appears on the visual form is the distorted version of the data. Thus, in order to reduce the gap, two objectives have been set. The first objective is to identify the similarity among the techniques through their behaviors. The second objective is to analyze the impact of such behaviors to visualization awareness. To achieve the objectives, fifteen commonly dot overlapped management techniques have been identified and their behaviors are analyzed. The behaviors are then grouped into six main behaviors. Finally, ten impacts of the behaviors to visualization awareness have been identified. Such awareness is important as part of the input during pattern extracting stage.

**Keywords:** *Overplotting, Overlap, Data Visualization, Dot Management Technique*

## 1. INTRODUCTION

Data visualization is a mechanism used to visualize and analyse complex data in various fields of knowledge. It is a mechanism that assists people in understanding the significance in data through visual context. Data visualization is an art and skill of presenting data in a visual manner as such information contained in the data becomes apparent [1]. Using data visualization, patterns, trends and correlation are easier to recognize, expose and successfully be applied in the area of descriptive statistic, which are undetected in text-based data.

Data visualization are divided into several commonly used techniques such as maps, temporal, multidimensional, network and hierarchical. However, this research focuses on dot-plot based technique. The dot-plot technique is commonly used in scattered plot and revisualization [2]. A dot

represents data relative to location on map and graph.

In dot-plot based visualization, volume of dots plays a vital role where they indicate concentration of values relative to a giving location on either graph or maps. Fewer dots indicate lower concentration while more dots indicate higher concentration for the particular location [3].

The dots can be in a single or multiple colours as well as various sizes. Each dot represents a phenomenon or a group of phenomena and their combination create phenomena density. The density reveals data patterns, which are knowledge hidden within the data. Figure 1 depicts the process flow of revisualization.



Figure 1: Process flow of geovisualization [4].

However, when utilizing dot-plot technique, over plotting issues consistently occur especially in big data [5]. The dots overlapped on top of each other, which create convoluted and confusing visual [6] thus, data presented are not accurate. Hence, this overlapped dots need to be managed. There are various techniques for the purpose and each supports the management in its own unique manner.

Most researches analyze the techniques of overlapped management in isolation while grouping them into their similar features are rarely discussed. Moreover, the discussion on the features' impacts to the visualization awareness is not common.

Visualization awareness is where users are aware on extra information regarding the plotted data. Users must alert that plotted data that appear on graphs have gone through some modification and adjustment. For instance, some data are not plotted to their exact location on map. This is due to the fact that the data are overlapped on top of each other and since tabulation of data frequency is vital, the overlapped data are shifted to next available location. In such cases, the exact locations of data are not stressed as compare to frequency of the tabulation. A set of data might only be 87% correctly position on map since the plot is at their estimation locations. Another example, user must aware that the plotted data represent only a few percentage of the whole population. Moreover, some outliers have been removed from the visualization and users must aware of this during knowledge extraction.

Having this awareness, it becomes extra information during data exploration via data visualization technique. This awareness will influence the knowledge extracting stage as depicted in Figure 2.



Figure 2: Pattern and awareness as an input to knowledge extraction stage.

Thus, this paper is written with two objectives. First is to analyse the behaviours of overlapped dot management techniques and second, to identify the impact of the behavior to visualization awareness.

To achieve the objectives, the next sections of this paper discuss the background study of data visualization and overlapping dots in various scenarios. Then, the discussion continues with method section where dot overlapping management techniques are analysed. The techniques are then group into their similar behaviours. Finally, the impact of such behaviours on visualization awareness is discussed before a conclusion remark.

## 2. BACKGROUND STUDY

Nowadays, it is common for data scientist to use data visualization techniques to provide explore insight and evidence in data for decision making involving various sectors such as business, government and academic.

Data visualization is important for data exploration as a mean to extract knowledge hidden within the data through data patterns [7]. It also provides methods that allow users to evaluate important characteristics of the data such as distribution, range, shape, multimodality, variability, correlations between data and outliers. Moreover, data visualization is the best tools to present data of two or more groups from different entities in a single visual form where various data patterns may derive different conclusions.

There are many data visualization techniques used to plot data. For instance, maps (choropleth, cartogram, dot distribution, proportional symbol), temporal (connected scatter plot, time series, Gantt chart, stream graph, Sankey diagram), multidimensional (pie chart, histogram, scatter plot, bar chart, bubble chart), hierarchical (dendrogram, tree map, sunburst diagram, circle packing) and network.

One of the widely used visualization techniques is dot-plot [5]. Its strength lies in its ability to show data distribution within certain region, hence, it can be a good starting point to get to know a distribution [8]. Dot plot is commonly use in scattered plot, box plots, violin plots as well as map plots where individual data points and summary statistics data are represent in overlaid on these visual forms [2].

Scatterplots is one of the most widely used graphs to present statistical data [9, 10, 11]. The graph visualizes two continuous variables using visual marks mapped to a two-dimensional Cartesian space. Dots of varies color, size, and shape mark the data distribution on the graph. In addition, scattered dot-plot is useful for displaying large sets of discrete values, especially when users not desire to start the quantitative scale at zero, which is required whenever bars are used [12]. It is easier to see the differences in data as compare to using bars that extended from a base value of zero. Dot plots can accommodate both positive and negative values simultaneously. Hence, it allows one to easily compared values of a continuous variable across different study ranges.

However, in huge dataset [14] individual dot may begin to overlap on top of each other. Parts of the reasons are due to larger dots and insufficient screen space to fit all the dots at the desired resolution. Another reason is that several data share the same value, which ended them on the same point on the graph. As a result, multiple or various dots are plotted on top of each other. The scenario causes the real number of dots to visually obscure. Moreover, important visual signal such as color become partially or completely restricted, thereby reducing search efficiency [15, 16]. In addition, the overlapped dots distort the data patterns and resulting in inaccurate data frequency [4].

Overlapped dots can also lead to incorrect impression and interpretation on data patterns. For example, based on the Figure 3 [16], dots are plotting as blue in A and as red in B, it is difficult to extract the data patterns because the two categories are overlap with the same color. On the other hand, the plot C and D are plots in the form of overlap involving two different colors. Similarly, analysis is difficult because the occlusion of data by other data. According to [16, 17] over plotting is a problem not just for scatterplots, but also for curve plots, 3D surface plots, 3D bar graphs and any other plot type where data can be occluded. The overlapped dots are difficult to manage because it depends on the number of data points and difficult to predict prior to visualization plotting.

Efforts have been taken to solve the issue of overlapping dots. Various techniques have been proposed as a result to manage the issue. Since these techniques modify the dots in various manners such as location, size, color and quantity, they definitely give impact to the data visualization. Users must aware of such impact and it is call visualization awareness.

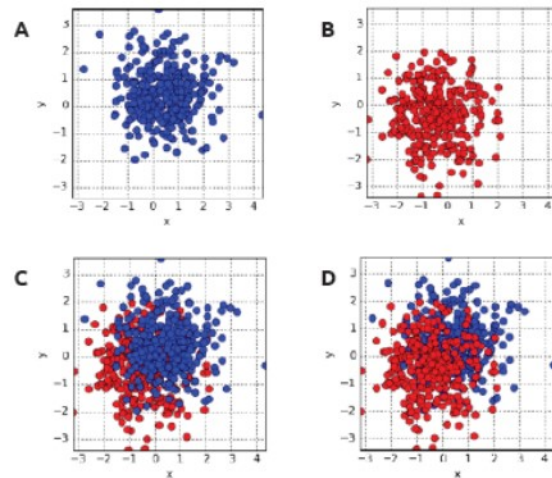


Figure 3: Overplotting [16].

Based on literature, data awareness is very important to ensure the data is correct, accurate and reliable. The purpose of the data is to provide information to help make decisions [13]. Data awareness can assist users to select data sets that provide accurate answers to the questions they are attempting to answer or one that meet the objectives [13]. Before actively looking at data sources, it is necessary to define the needs of the data. Being able to specify what it is you are trying to find out or what you are hoping to achieve from the outset, will help ensure you get the data you require to make well informed decisions [13]. Then, in order to correctly visualize the data, visualization awareness will help the user to identify the suitable visualization techniques to represent effectively the insight of the data.

Most of researches reviewed the visualization techniques before mapped them to their suitable applications and data. However, the techniques are not group into their common behaviors and their impact to the visualization awareness such as overlapped dots is not commonly studied.

Thus, this research has been conducted due to the identified gaps by identifying the behaviors of the most commonly overlapped dot management techniques and the impacts of such behaviors to visualization awareness. The finding is vital to assist users during knowledge extraction stage, as the users are aware of the data state. Eventually it establishes guidance to researches. Next section discusses method to achieve the two research objectives.

### 3. METHOD

In order to group the overlapped dot management techniques into their similar behaviours and to discover the impact of such behaviour to visualization awareness, a few steps have been taken. First, fifteen commonly used techniques in various scenarios are studied. Then, the main behaviours of the techniques are identified and grouped. Finally, the impacts of the behaviors on visualization awareness are analysed. The following sections discuss the steps in further details.

#### 3.1 Techniques of Distribution in Overlapped Dots

There are many approaches available to manage overlapped dots. The approaches used for nonlinear optimization technique is dot pattern generation task. Its runs faster as compare to existing simulation based methods and it generates dot patterns with comparable quality. The process starts by generating an initial dot patterns from a dot-density and simulation based method. Then, all dots are replaced within circles at an appropriate radius computed from the density. Lastly, all the dots are placed and put at the right positions in the circles [18]. It can reduce complex calculation cost and produce a better dot quality diffuser pattern on edge.

Besides, pixel-oriented visualization is another technique used to solve the overlap. The first method called nearest-neighboring algorithm. The overlap points are shifted to the nearest unoccupied position. For example, Curve-based algorithm which is one to the nearest-neighboring group of algorithm, moves overlapping dots to unoccupied position in the direction of screen-filling curve which is similar to Hilbert-curve and Z-curve.

Grid-fit algorithm is a technique that separates the overlapped dots in a hierarchy manner. This algorithm adopts the concept of quad tree data structured. Each layer has its own data space in which it is divide into four sub regions. The division process ensures that the region for the sub region is larger than the number of pixels it fits. Based on [19] the spatial distribution becomes more intuitive to the sediment movement.

Another technique is circle method. The method generates an initial random dot pattern using a randomized LDS and remove unwanted inter-dot overlaps in the initial pattern. For multi-sphere schema, simulation based approach such as L-BFGS is suitable because it is faster and improve the quality of the dot patterns especially near the edges.

In generally, the Self-Organizing Map (SOM) is a tool for visualization high dimensional data and successfully applies to visualizing spatial data (GIS application) [20]. SOM cluster geo-referenced data before integrates them with visualization technique. SOM uses visualize spatial data as well as colors to link between the output space of the network and geographical visualization. This approach suitable for low dimensional surface and is effective for clustering. It also uses colors for representing distances between references vectors associated with the SOM units. When using SOM technique, it shows increases on the quality of the analysis but it is difficult to classify correctly the geo-referenced element [20].

Image processing is another method used to redistribute the overlapped image data. For instance, to count the population of birds that migrates to other continent of the globe. The technique involves four main steps in the process of redistributing the overlapped data [21]. Firstly, noisy dots and background are removed in the bird image using multilevel threshold. Secondly, seek the bird shape by finding area value to segment the image. Thirdly, the distance transform process identifies the distance between images. Finally, the erosion technique is use for object reduction, which identifies pixel relationship in four directions.

Histogram equalization also used the technique of image processing to increase dynamic distance and contrast image [22, 23]. The benefit of using this technique is that it can increase the performance of identifying bird population. This process compares a given pixels of histogram to share with each gray value to the reference histogram and calculate the equation production. This concept uses the map dot for emphasize density variation [24]. The gradient of color scale serves as a reference histogram. So that, the color assigned to each pixel on the map becomes proportional to the reference distribution used to display data values.



Another technique is using line visualization to manage overlapped dots. Line is simplest and easiest to represent the data visualization in visual form. In addition to data display, line visualization support data exploration and analysis. Each data is presented as lines plot. Figure 4 displays four different line algorithms in an attempt to redistribute data frequency [4]. Some of the line version are straight line, cross and star to redistribute the overlapped dots.

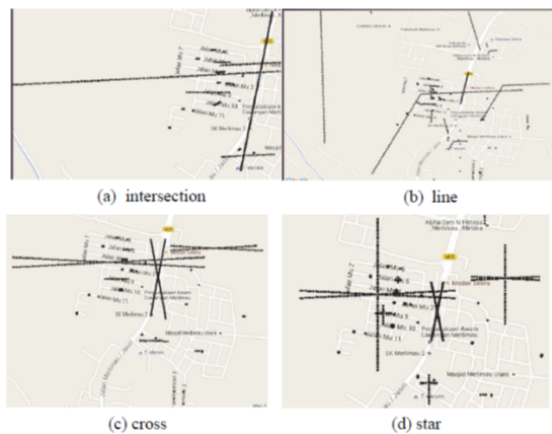


Figure 4: Four line algorithms to display data density [4].

Splatter plot is one of the techniques to solve the overlapped problem in data visualization. This technique is new for displaying scattered data which remains interpretable, as the number of points grows large. The based key of this technique is explicitly performing abstraction to reduce visual clutter and perform abstraction in screen space [25]. The concept of Splatterplots is to maintain a readable display, regardless of zoom level [25]. Figure 5 show the difference between Splatterplot and scatter plot with zoom level increasing from left to right. Splatter processes each subgroup of data separately to produce dense region, rendering dots in sparse regions of the map in the right location to indicate outliers [25] and then combine them into single splatterplot. The advantage of splatterplot is that its displayed size is small but its manages to reveal general trends in the data. In addition, splatterplot also support users to make judgement on the data when dealing with small plots. Splatterplots retains ability to see the overall shapes, trends, relationships between set and a sense of outlier although the number of data points greatly exceeds number of pixels.

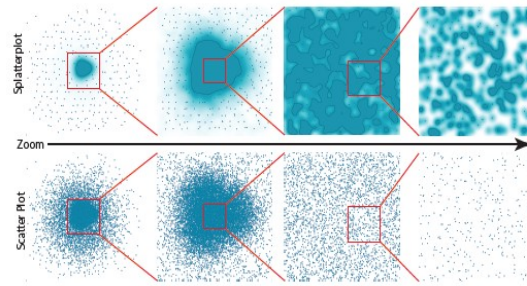


Figure 5: Effect of zooming on detail in a Splatterplot and scatter plot [25].

A researcher by [27] proposes a stacking technique to provide a view of dense data that reveals a range of magnitude considerably larger in color and representations area. Stacking methods can be in two dimensions that are 2D and 3D. Stacking 3D plot of a manifold can reveal mathematical structure which cannot be seen in a 2D contour plot. Thus, stacking can reveal a database structure not evident in 2D panels or contours. Stacking is more appropriate when other colors, areas or other aesthetic variables have no dynamic distance to allow for extreme magnitude comparisons. It is not limited to points and lines.

Some researchers use clustering method such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN). DBSCAN is a spectral clustering method that can tolerate noise in data [28]. Spectral clustering methods, on the other hand are quality cluster production and can easily implicate overlap of dots. It is also being designed for arbitrarily clusters shape without posting any constrains on the cluster. However, the number of clusters must be defined before implementation [29].

Mean-shift is a non-parametric clustering technique [30]. This method used to detect the probability of distribution from the set of discrete samples to identify local maxima and area of associated to the maxima [30]. Therefore, a mean shift is used as an algorithm to locate the local maximum (mod) as well as clustering techniques (mod-related areas). Mean-shift algorithm run for a set of different initial positions in order to identify all local maxima.

Pixel Maps is a new pixel-oriented visual data mining technique for large spatial datasets [31]. It combines two techniques that are kernel density-based clustering and novel pixel-based displays to emphasize clusters while avoiding overlap in locally dense point sets on maps. It provides every data point to a unique pixel in 2D

screen space, and the tradeoff tray from spatial location (maintenance of absolute and relative positions). It optimizes the problem and practical value for exploring geo-spatial statistical data. Pixel Map is an example of distortion technique in data visualization, which expands the dense region of a map in order to achieve the most suitable dots and histoscale [31, 32]. It scales the map along two dimensions of Euclidian in which distortion degrees depend on region in a particular the number of dots.

Other technique is bubble plot. A bubble plot is a variation of a scatter plot in which the markers are replaced with bubbles [26]. A bubble plot displays the relationships among at least three measures. Each bubble represents an observation. Mostly bubble plot is useful for data sets with dozens to hundreds of value or when the values differ by several orders of magnitude. Some bubble plot technique use color to represent the overlapped on a geographical map. The so-called "bubble plot" was created by Playfair [33] and has been used for more than two centuries.

Block-structured adaptive mesh refinement (AMR) is used to resolve fine scale features in the flow such as shocks and detonations [42]. This method adopts the accuracy of a solution within certain sensitive region of simulation, dynamically [42]. It is used to represent moving boundaries, and these grids overlap with stationary background Cartesian grids. Refinement grids are added to base level grids according to an estimate of the error, and these refinement grids move with their corresponding base-level grids. A new set of refinement grids are generated to cover all flagged point and the solution is transferred from the old grid hierarchy to the new one. Since the regridding procedure takes place at a fixed time, it is effectively decoupled from the moving grid stage. Figure 6, shows the movement overlap grid with AMR [42].

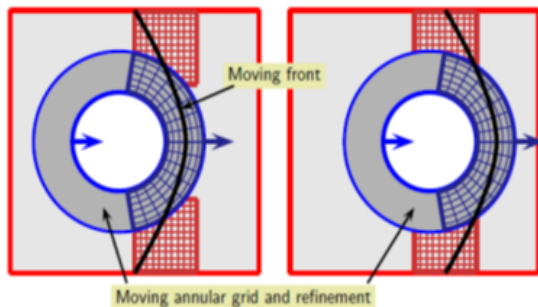


Figure 6: A moving overlapping grid with AMR [42].

### 3.2 The Behaviors of Dot Management

The overlapped dot management technique has been analyzed and their main behaviors have been identified. This study groups those techniques into six categories based on the behaviors of the techniques. The six behaviors are jittering, refinement, distortion, aggregation, density, and clustering as depicted in Figure 7.



Figure 7: Grouping techniques into their similar behaviours.

#### i. Jittering

Jittering reduces the number of overlapping points by replacing it to other available location [22]. This has the effect of spreading the dots so they are more easily distinguished. The process of identifying the new possible location can be at random or using predefined grid method. The exact location of the dot is not important since the stress is on data frequency within certain region. The distribution is flat on single layer where the view is more toward the whole distribution of data.

The issue of jittering is that the arrangement of point space on the result map tends to appear arbitrary and does not represent real data. There are times the dots ended on illogical location such as on river. However, it can be controlled by introducing a restricted boundary that limits how

the dots are spread or by producing more natural dispersion patterns that put dots in false layouts [34, 35, 36].

The size of the dots must be select wisely. Too large symbol on huge data will lead to over exaggerating the patterns density. However, too small dots lead to extreme distance between the dots that impose a feeling of under density [17].

However, most jittering mechanisms apply for arbitrary shifts to all data points, irrespective of whether they overlap or not. These have the disadvantages of disturbing all points from their actual location on the visual canvas, and still do not completely over plotting.

This is a useful method of representing bootstrap estimates in a public space [37]. For a smaller dataset, this approach works well. However, the plotting area will quickly downgrade as N increased. Consequently, jitter is suitable to limited applications.

## ii. Refinement

Refinement technique is another technique to address visual clutter by reducing the number of dots on the map [22]. Density preserving algorithms [38, 39] used refinement to obtain subset of dot density in the actual data. For example, in considering positioning of deer, the technique group around deer showing the second order relationship [38]. However, if some physical features in the landscape separate the deer then the grouping may have no biological meaning and therefore irrelevant. Otherwise, thematic information is use as a subset selection. Refinement technique includes separation, merger and transfer of dots and reduces the shape and size of the overlap dot.

This creates a simplified dot that retains data-specific features. The refinement reduces the overlap and maintains the spacing of the dots. Moreover, the relative density differences in the actual data, reduces the map delimiter as it eliminates potentially relevant information. Refinement techniques does not change the geographical position of the dots, thus a large amount of continuous overlap between the remaining points and density differences between regions of the map have been preserved. Furthermore, when using refine, overlap is not address and most dots cannot be taken consideration during analysis.

## iii. Distortion

Distortion behavior change the space arrangement of dotted spaces in the dense area of the map so that each dot can be displayed at a unique position (ie non-overlapping) [22]. Distortion technique gains profits by revealing delicate details in the data. However, they tend to lose in terms of maintaining visual references to known geographical features. For example, roads, rivers, and administrative borders can become unrecognizable, while excessive shrinking can make the area rarely read. Additionally, the available display space on the map limits the extent to which effective distortion is to reduce overlap.

Challenges imposed on a distorted map are to balance between reducing the number of overlapping and horizontal dots of the map until the geographical feature remains distinct.

## iv. Aggregation

Aggregation simplifies dots map so that different density patterns are more distinguished from each other [22]. Once a conceptual scenic object has been determined, it can be used to guide the aggregation of data into a fixed-size grid [16]. All supported aggregate options are now implement as an additional reduction operator. Given aggregate bin for updating, it usually corresponds to one final pixel as well as a new data point. The reducing operator updates the bin state in several ways. Figure 8 below shows the visualization of various aggregations using datashader.

Unfortunately, the size symbol for representing aggregates improves but does not eliminate overlap; bubbles or other elements of representation are still close to one another when its size exceeds nearest-neighbor distances. Furthermore, using area to represent magnitude incurs nonlinear distortions in perception [8, 32, 37].

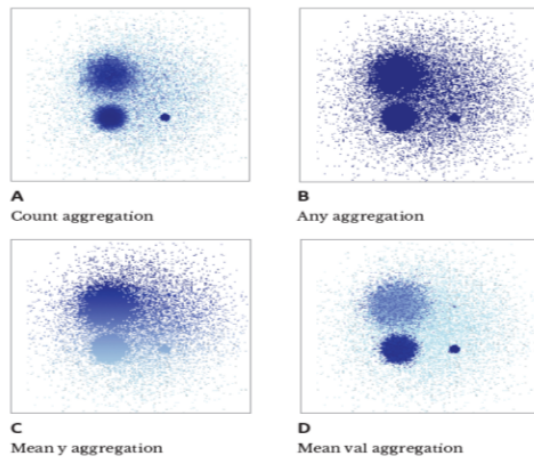


Figure 8: Visualization of various aggregations using datashader [16].

#### v. Density

The fourth technique to manage overlap is density where color plays a vital role. Density is a common way of dealing with excessive issues, as they remove the discrete glyph idea for data points [25]. Density is an important metric because it indicates stronger relationships between points within a cluster. However, most implementation of density follow rules of thumb or statistical models to compute the ideal amount of smoothing in data space. Density a simple way to direct tie amount of smoothing used to create density functions directly to screen space. For data visualization, data density is directly related to accuracy of the visual produced.

Density can be divided into depth dimension and flat. A depth dimensions is like 3D visualization. The dots use coding and two colors blending together making it easy to see solid data displayed bundling with multiple subgroups in interactive systems. Figure 9, depicts details about the density through colors. Solid areas are aggregate into seamless contours. While flat dimensions are areas in which dots are shaded, color, or patterned according to the value of a given attribute. Colors can be in the form of thematic or shaded [40]. Shaded areas where darker areas represent higher concentration of a given data. It is important to choose the right colors for dot presentations as they can affect how one interprets the information presented especially the data density. Similar-size dots are recommended to the extent possible, as a few larger dots can dominate a map, leading to miss interpretation of information.

The impact when using density for solving the overlap is that the original data point does not show in density displayed, which is important for doing direct queries of the data when the specific values of points of interest are important.

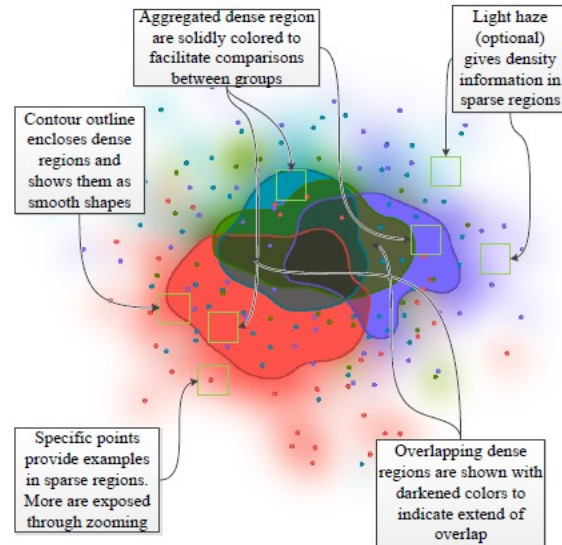


Figure 9: Splatterplot with explanation [25].

#### iv. Grouping

Group is defined to have homogeneity within clusters and high heterogeneity among clusters. This is accomplishing by grouping unit that are similar according to some appropriate distance criterion. The choice of the distance criterion plays therefore a leading role in the group's definition [41]. Grouping is the process of grouping physical or abstract objects into classes of similar object [R29].

The impact when using the grouping behaviour is that it is easier to determine the dots position among other dots such as between group, within group as well as between and within group cluster. This simple technique is frequently used to analyse clustered data in smaller data visualization.

Figure 10, summarized the behaviours of the 15 techniques and the impact of such behaviours to the visualization awareness.



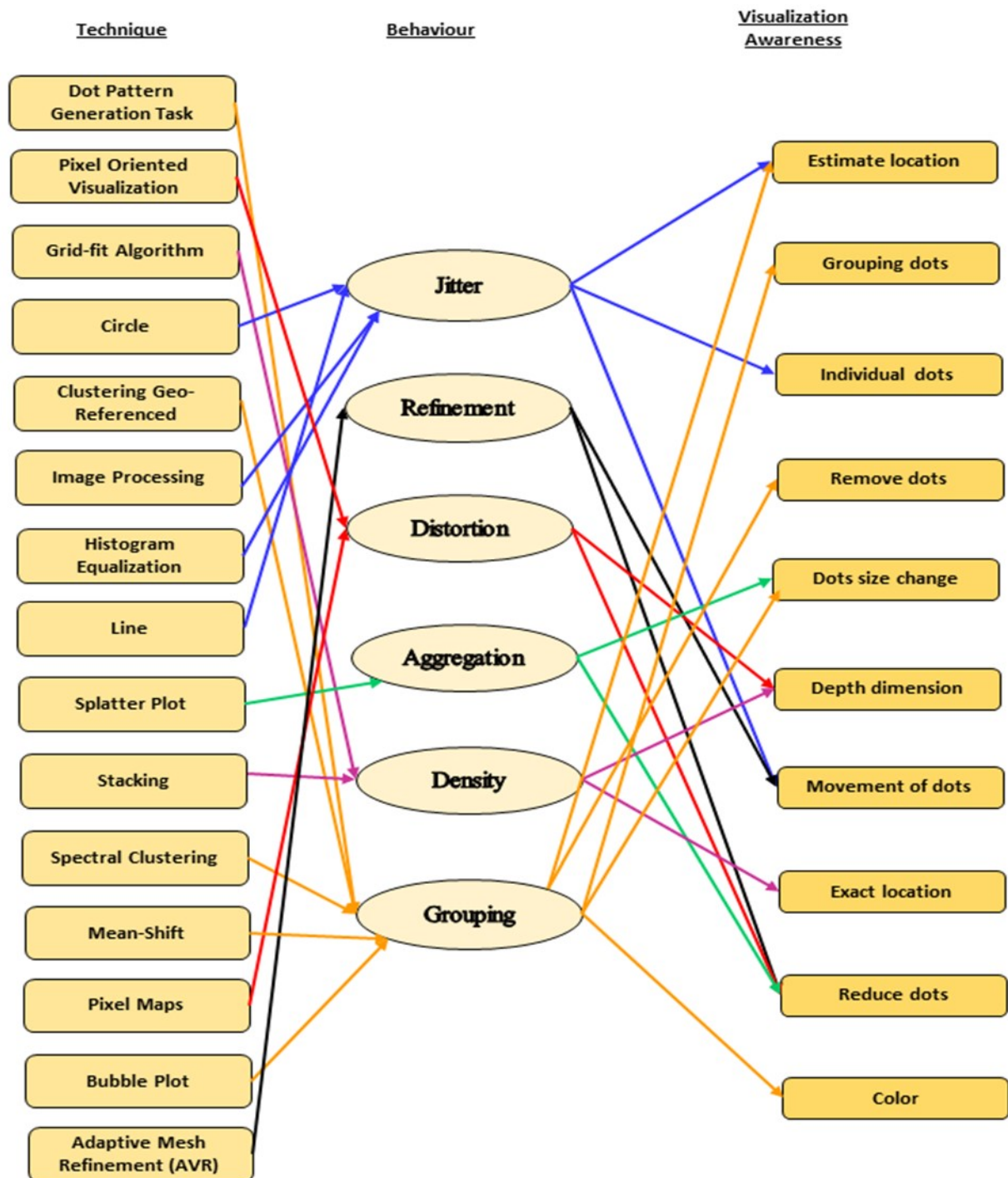


Figure 10: Behaviors And Visualization Awareness Of Overlapped Dot Management Techniques

#### 4. CONCLUSIONS

In summary, this paper analyzes the relationship between the behaviors of dot

overlapped management techniques and their impacts to the visualization awareness which was usually being ignored. Awareness is where users understand and aware that the data has gone

through some modification prior to plotting them on visual form. The study starts by taking a step to understanding the behavior of fifteen techniques used to solve overlapped dot problems in visualization, which involves small or large number of data. Then, these techniques are categorized into six group behaviors which are jittering, refinement, distortion, aggregation, density and clustering. Finally, the behaviors are matched to their visualization awareness which are grouping dots, individual dots, reduce dot, dot size change, depth dimension, movement dot, exact location, estimate location, remove dot (outliner/ frequency) and color. Results demonstrations the relationship patterns of the techniques and the behaviours impacts to the visualization awareness. Thus, the findings indicate that the identification of the techniques similarities is vital and contribute to the knowledge extraction and visual patterns. For good measure, it became part of the input which is essentially need to be considered. Subsequently it provides guidance which is significant to researches.

## ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknologi MARA, Malaysia for the facilities and financial support under the LESTARI grant 600-IRMI/Dana KCM 5/3/LESTARI (158/2017).

## REFERENCES:

- [1] M. Negyal, "Data Visualisation 12," Chart, no. May, pp. 1–12, 2012.
- [2] T. L. Weissgerber, M. Savic, S. J. Winham, D. Stanisavljevic, V. D. Garovic, and N. M. Milic, "Data Visualization, Bar Naked: A Free Tool for Creating Interactive Graphics," J. Biol. Chem., p. jbc.000147.2017, 2017.
- [3] <http://www.caliper.com/glossary/what-is-a-dot-density-map.htm>
- [4] Z. Idrus, S. Z.Z.Abidin, N.Omar, Z. Idrus, and N. S. A. M. Sofee, "Geovisualization Of Non-Resident Students' Tabulation Using Line Clustering," Reg. Conf. Sci. Technol. Soc. Sci., 2016.
- [5] T. Bogon, F. Lorig, and I. J. Timm, "Visualizing the impact of probability distributions on particle swarm optimization," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2013, vol. 7928 LNCS, no. PART 1, pp. 120–128.
- [6] W. S. Cleveland et al. "The elements of graphing data", Wadsworth Advanced Books and Software Monterey, 1985.
- [7] E. Z. Martinez, "Description of continuous data using bar graphs: A misleading approach," Rev. Soc. Bras. Med. Trop., vol. 48, no. 4, pp. 494–497, 2015.
- [8] M. C. Minnotte, S. R. Sain, and D. Scott, "Multivariate visualization by density estimation". In Handbook of Data Visualization, pages 389–413. Springer, 2008.
- [9] W. S. Cleveland and M. E. McGill, "Dynamic graphics for statistics", CRC Press, 1988.
- [10] N. Elmquist, P. Dragicevic, and J.-D. Fekete, "Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation", IEEE Transactions on Visualization and Computer Graphics, 14(6):1539–1148, 2008.
- [11] J. M. Utts, "Seeing Through Statistics", Duxbury Press, 1996.
- [12] S. Ekonomiczne, Z. Naukowe, and G. Ko, "Data Visualization In The Resampling," no. 247, 2015.
- [13] Australian Bureau of Statistics, 1500.0 – A guide for using statistic for evidence based policy, 2010 . Canberra: ABS, 2010. [Online]. Available from AusStats, <http://www.abs.gov.au/ausstats/abs@.nsf/lookup/p1500.0chapter62010>. [Accessed: Okt. 7, 2017].
- [14] M. J. Bravo and H. Farid, "Recognizing and segmenting objects in clutter," Vision Res., vol. 44, no. 4, pp. 385–396, 2004.
- [15] M. J. Bravo and H. Farid, "Search for a category target in clutter," Perception, vol. 33, no. 6, pp. 643–652, 2004.
- [16] J. Bednar, "Big Data Visualization with Datashader," no. August, 2016.
- [17] D. Park, S.-H. Kim, and N. Elmquist, "Gatherplots: Generalized Scatterplots for Nominal Data," 2017.
- [18] T. Imamichi, H. Numata, H. Mizuta, and T. Id?, "Nonlinear optimization to generate non-overlapping random dot patterns," in Proceedings - Winter Simulation Conference, 2011, pp. 2414–2425.
- [19] K. Forsythe, C. Marvin, C. Valancius, J. Watt, J. Aversa, S. Swales, D. Jakubek, and R. Shaker, "Geovisualization of Mercury

- Contamination in Lake St. Clair Sediments,” *J. Mar. Sci. Eng.*, vol. 4, no. 1, p. 19, 2016.
- [20] J. Gorricha and V. Lobo, “Improvements on the visualization of clusters in geo-referenced data using Self-Organizing Maps,” *Comput. Geosci.*, vol. 43, pp. 177–186, 2012.
- [21] N. Tidmerng, W. Songpan, and M. Wattana, “Solving bird image overlapping for automatic population counts of birds using image processing,” 2016 *Manag. Innov. Technol. Int. Conf. MITiCON 2016*, vol. 1, no. 1, p. MIT84-MIT87, 2017.
- [22] A. Chua and A. Vande Moere, “BinSq: visualizing geographic dot density patterns with gridded maps,” *Cartogr. Geogr. Inf. Sci.*, vol. 44, no. 5, pp. 390–409, 2017.
- [23] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, “Adaptive Histogram Equalization And Its Variations,” *Comput. vision, Graph. image Process.*, vol. 39, no. 3, pp. 355–368, 1987.
- [24] E. Bertini, A. Di Girolamo, and G. Santucci, “See What You Know: Analyzing Data Distribution to Improve Density Map Visualization,” in *Eurographics IEEEVGTC Symposium on Visualization*, 2007, pp. 1–7.
- [25] A. Mayorga and M. Gleicher, “Splatterplots: Overcoming overdraw in scatter plots,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 9, pp. 1526–1538, 2013.
- [26] Sas, “Data Visualization Techniques,” *White Pap.*, pp. 2–16, 2013.
- [27] T. N. Dang, L. Wilkinson, and A. Anand, “Stacking graphic elements to avoid overplotting,” *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1044–1052, 2010.
- [28] D. Villatoro, J. Serna, V. Rodríguez, and M. Torrent-Moreno, “The TweetBeat of the city: Microblogging used for discovering behavioural patterns during the MWC2012,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7685 LNAI, pp. 43–56.
- [29] R. Rösler and T. Liebig, “Using data from location based social networks for urban activity clustering,” in *Geographic Information Science at the Heart of Europe*, 2013, pp. 225–245.
- [30] V. Frias-Martinez, V. Soto, H. Hohwald, and E. Frias-Martinez, “Characterizing urban landscapes using geolocated tweets,” in *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, 2012, pp. 239–248.
- [31] D. A. Keim, C. Panse, M. Sips, and S. C. North, “PixelMaps: A New Visual Data Mining Approach for Analyzing Large Spatial Data Sets,” in *Third IEEE International Conference on Data Mining*, 2003, vol. 1, pp. 565–568.
- [32] D. A. Keim, C. Panse, M. Schafer, M. Sips, and S. C. North. 2003b. “Histoscale: An Efficient Approach for Computing Pseudo-Cartograms.” in *Proceedings of the 14th IEEE Visualization Conference (VIS’03)*, 93. New York: IEEE. [http://www.computer.org/csdl/proceedings/ieee\\_vis/2003/2030/00/20300093.pdf](http://www.computer.org/csdl/proceedings/ieee_vis/2003/2030/00/20300093.pdf), 2003.
- [33] W. Playfair, “The commercial and political atlas and statistical breviary”, Original edition (1786) edited and republished by H. Wainer and I. Spence. Cambridge University Press, Cambridge, 1786.
- [34] A. J. Kimerling, “Dotting the Dot Map, Revisited,” *Cartogr. Geogr. Inf. Sci.*, vol. 36, no. 2, pp. 165–182, 2009.
- [35] A. Hey, “Automated Dot Mapping: How to Dot the Dot Map.” *Cartography and Geographic Information Science* 39 (1): 17–29. doi:10.1559/1523040639117, 2012.
- [36] A. Hey, and R. Bill, “Placing Dots in Dot Maps.” *International Journal of Geographical Information Science* 28 (12): 2417–2434. doi:10.1080/13658816.2014.928822, 2014.
- [37] L. Wilkinson, “The Grammar of Graphics”, Springer-Verlag, New York, 2nd edition, 2005.
- [38] D. Burghardt, R. S. Purves, and A. J. Edwardes, “Techniques for on The-Fly Generalisation of Thematic Point Data Using Hierarchical Data Structures.” in *Proceedings of the 12th Annual Conference on GIS Research UK*, Norwich, April 28–30, 2004.
- [39] S. Peters, “Quadtree and Octree Based Approach for Point Data Selection in 2D or 3D.” *Annals of GIS* 19 (1):37–44. doi:10.1080/19475683.2012.758171, 2013.
- [40] S. Yasobant, K. Suresh Vora, C. Hughes, A. Upadhyay, and D. V. Mavalankar, “Geovisualization: A Newer GIS Technology for Implementation Research in Health,” *J. Geogr. Inf. Syst.*, vol. 7, no. 7, pp. 20–28, 2015.

- [41] H.-M. Wu, S. Tzeng, and C. Chen, “Handbook of Data Visualization,” Handb. Data Vis., no. Chapter 8, pp. 681–708, 2008.
- [42] W. D. Henshaw, “ScienceDirect - Journal of Computational Physics: Moving overlapping grids with adaptive mesh refinement for high-speed reactive and non-reactive flow,” no. September 2005, pp. 1–44.